Deep Reinforcement Learning: Theory and Application

Misha Obukhov, UCSB Computer Engineering

Mentor: Mert Torun   Faculty Advisor: Dr. Ramtin Pedarsani

Introduction

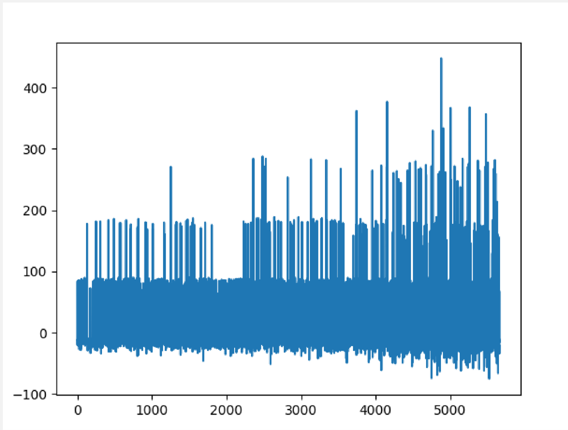
How can we create something that learns how to act in its environment?

Here, we apply the Proximal Policy Optimization (PPO 2017) Reinforcement Learning algorithm to solve a simple environment, and demonstrate multiagent convergence in a zero sum task.

Methods

1) Map Environmental State to data representation, including reward signal for changes between states.
2) Neural Network with 256 Neurons maps states to actions and determines state value.
3) PPO is used to train the network in batches of 20 step action/state chains.
4) For multiagent learning, an evolutionary algorithm is used to select the superior model and force it to train against itself.

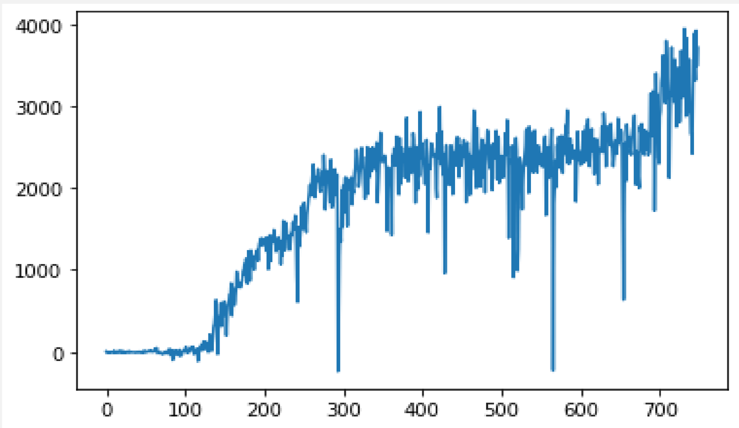
Results



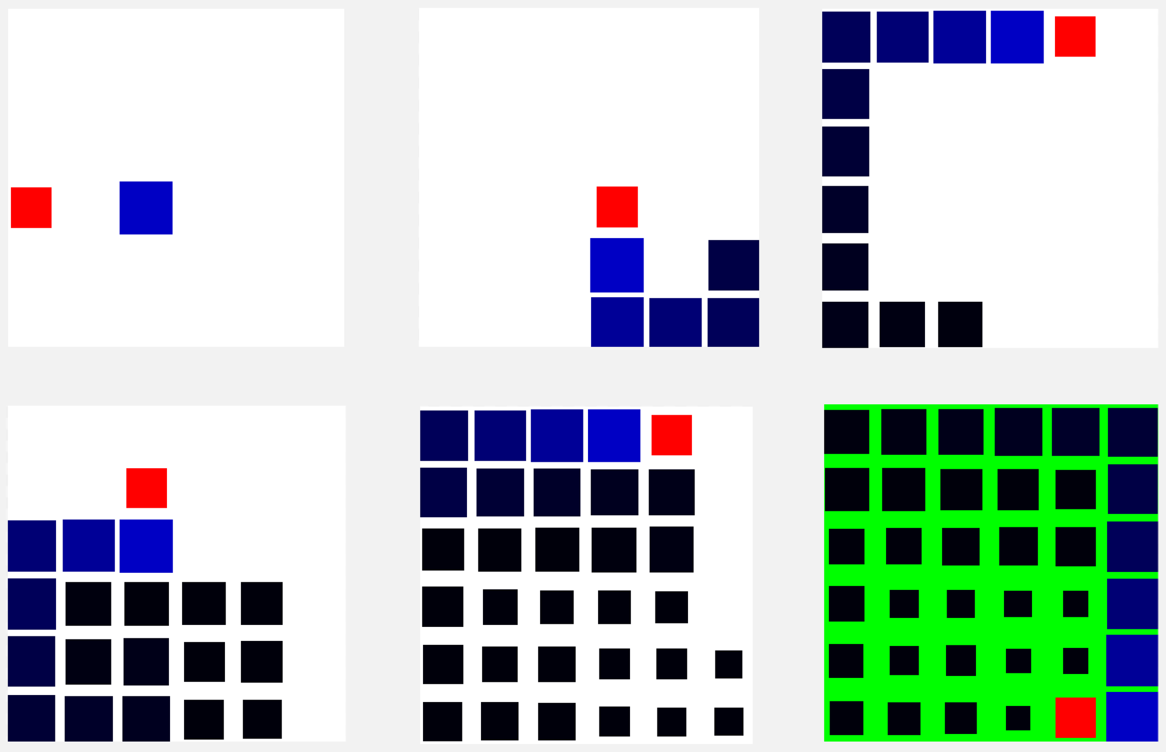
(100k Timesteps)	100 Neurons	30 Neurons
Training Average Reward	7.7	14.0
Testing Average Reward	15.75	31.8

Tests with OpenAI PPO1 template algorithm were unsuccessful, but better performance was achieved with less neurons.

Tests with PyTorch barebones PPO implementation proved convergence.



Algorithm was able to learn and consistently solve single agent game, after only 10k training games; around 10 minutes on i7 laptop.



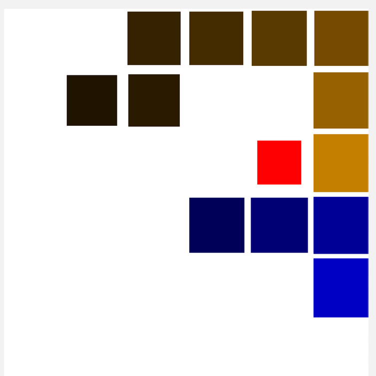
Multiagent learning in zero sum, competitive environment proved to achieve convergence and avoidance strategies.

Conclusions

Initial tests showed that less neurons would lead to better performance – probably due to increased capability for generalization.

PyTorch PPO implementation proved to show convergence after around 1000 games played.

Constant negative reward was necessary to teach snake to take efficient pathing.



Multiagent training was successful and particularly effective in the start of the process, as the snake must learn to avoid body more quickly.

Literature Cited

Schulman, J. 2017. Proximal Policy Optimization Algorithms. *arXiv:1707.06347*.

Acknowledgments

CSEP and Eureka Program

Further Information

mobukhov@ucsb.edu